

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Apparent Paradoxes in Moral Reasoning; Or how you forced him to do it, even though he wasn't forced to do it.

Permalink

<https://escholarship.org/uc/item/6q04z6jf>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN

1069-7977

Authors

Phillips, Jonathan
Young, Liane

Publication Date

2011

Peer reviewed

Apparent Paradoxes in Moral Reasoning; Or how you forced him to do it, even though he wasn't forced to do it.

Jonathan Phillips (jonathan.phillips@yale.edu)

Department of Philosophy, Yale University
New Haven, CT 06511 USA

Liane Young (liane.young@bc.edu)

Department of Psychology, Boston College
Chestnut Hill, MA 02467 USA

Abstract

The importance of situational constraint for moral evaluations is widely accepted in philosophy, psychology, and the law. However, recent work suggests that this relationship is actually bidirectional: moral evaluations can also influence our judgments of situational constraint. For example, if an agent is thought to have acted immorally rather than morally, that agent is often judged to have acted with greater freedom and under less situational constraint. Moreover, when considering interpersonal situations, we judge that an agent who forces another to act immorally (versus morally) uses more force. These two features can result in contradictory response patterns in which participants judge both that (1) a forcer forced a forcee to act and (2) the forcee was not forced by the forcer to act. Here, we characterize potential psychological mechanisms, in particular, “moral focus” and counterfactual reasoning that account for this paradoxical pattern of judgments.

Keywords: force; morality; counterfactual thinking; focus

When evaluating moral agents, we consider whether the agent acted freely or under situational constraint (Darley & Shultz, 1990; Nichols & Knobe, 2007; Woolfolk, Doris, & Darley, 2006). Both ordinary folk and legal scholars consider this a crucial factor for moral judgment (Hart, 1968). Emerging research, however, suggests that moral judgments in turn influence our evaluations of situational constraint (Phillips & Knobe 2009). That is, given a fixed level of situational constraint, people often judge that an agent is *not* forced to perform a morally impermissible action, though she *is* forced to perform a morally permissible action.

Consider an example originally proposed by Aristotle, featuring a sailor in the midst of a raging storm (*Nicomachean Ethics* 1110a10-11). The sailor, knowing that he must make his ship lighter to avoid capsizing, throws his expensive cargo overboard. When presented with this case, participants judge that the sailor was forced to throw his cargo overboard. However, changing the moral valence of the sailor's action changes judgments of force. If the sailor threw his wife overboard instead, participants judged

that the sailor really wasn't forced to do so, despite the fact that the situational constraint (the threat of capsizing) was the same (Phillips & Knobe 2009). Thus, moral judgments appear to alter judgments of situational constraint and *force*. While this effect may be interesting, the influence of moral judgments on force has turned out to be more complex.

In many moral situations, multiple agents may be responsible for a single bad outcome. For example, if the sailor was ordered by the captain of the ship to throw the passengers overboard, then both the captain and the sailor may be partially responsible for the death of the passengers. The combination of the fact that moral judgments influence judgments of situational constraint, and the fact that multiple agents can be morally responsible for a single outcome has the following consequence: shifting the primary focus from one moral agent to another within a single scenario leads observers to make contradicting judgments of situational constraint (Young & Phillips 2011). When observers focus not on the ship captain but on the *sailor* as the morally blameworthy agent, they judge that the sailor *freely* threw the passengers overboard, and that he was not truly forced by the ship captain. However, if observers redirect their focus to the *captain*, then they are more likely to agree that the captain *did* force the sailor to throw the passengers overboard.

We refer to this contradictory response pattern (i.e. X forced Y, but Y was not forced by X) as the paradox of moral focus. Our current aims are (1) to illustrate this same paradoxical pattern of responses in two new scenarios that may actually occur in ordinary, everyday life, and (2) to develop a model of the cognitive mechanisms that underlie this effect.

Notably, this research may be independently important for the recent debate on the nature of the influence of moral cognition on non-moral cognition (Guglielmo, Monroe, & Malle, in press; Machery, 2008; Mallon, 2008). This debate hinges on whether this influence reflects a legitimate role for morality in concepts such as force (Knobe, 2010; Pettit & Knobe, 2009; Phelan & Sarkissian, 2008; Wright & Bengson, 2009), or whether this influence merely reflects motivated moral reasoning (Alicke, 2008; Nadelhoffer, in press). Our contribution aims to help move this debate

forward by positing a specific underlying cognitive mechanism for the influence of morality on judgments of force.

We will begin by outlining this model for the paradox of moral focus and then present two studies that suggest (1) such effects may occur in ordinary, everyday situations and (2) that the mechanisms of *moral focus* and *counterfactual thinking* play fundamental roles in generating the paradoxical pattern of responses. We then discuss the past and current studies in terms of the proposed account of the underlying cognitive mechanisms. Specifically, we argue that this account can explain the influence of morality on putative non-moral judgments of force as well as the paradox of moral focus.

Theoretical Model

Philosophers have long understood that to be ‘forced’ entails lacking the ability to do otherwise. Thus Aquinas discussed ‘the necessity of coercion,’ by saying: ‘When someone is forced by some agent...he is not able to do the contrary’ (Summa Theologica I.II Q6 A6). This suggests that to claim that an agent was ‘forced’ is at least in part to say that the agent could not do anything other than what he did. Moreover, to say that an agent was *not* forced is to say he *could have* acted differently. If people’s intuitive judgments of force reflect this principle, either consciously or unconsciously, then their judgments of situational constraint are partially determined by the extent to which participants believe an agent could have acted differently.

Importantly, to imagine whether and how an agent could have acted differently is to engage in counterfactual reasoning. Past research has targeted the impact of moral judgment on counterfactual reasoning. In one such study, participants engaged in more counterfactual thinking for “immoral” versus neutral causes of an outcome (N’gbala & Branscombe, 1997). Thus, the influence of morality on judgments of whether an agent was forced to act seems to have a straightforward explanation: When participants are faced with an immoral action, they focus on the agent who performed the action and are more inclined to entertain alternative actions that the agent could have performed. This, in turn, leads participants to judge that the immoral agent acted freely – given the salient alternative options they perceive as available to the agent. Returning to the original study (Phillips & Knobe 2009), the hypothesis is that participants were more inclined to consider counterfactual alternatives to the sailor’s action when he threw his wife overboard than when he threw his expensive cargo overboard.

Yet, in situations in which more than one agent is responsible for a single immoral outcome, blame can be shifted from one agent to the other. Moral evaluations of different agents may motivate participants to focus on one agent versus the other. Consequently, the shift in moral focus may result in a corresponding shift in which counterfactual actions are considered – that of the *forcer* versus that of the *forcee*. This motivated reasoning should

therefore be consistent with participants’ moral judgments but, as a result, may reflect internal inconsistencies, as observed in focusing bias more generally (Legrenzi, et al., 1993). It is this process that we suggest leads to the paradox of moral focus.

Study 1

Study 1 presents a new instance of the paradoxical force judgments. However, the scenario used in this study, unlike scenarios used in earlier studies, is realistic enough to actually occur. If participants show the same inconsistent pattern of results in this more realistic scenario, we would have initial evidence that such paradoxical reasoning may influence situations similar to situations that have actually occurred.¹

All participants read about a military commander who ordered his subordinates to torture captured enemy soldiers for military information:

There was a military unit commander who wanted to help end a rebel uprising in his country. He decided that the best way to do this was to capture some of the rebel soldiers and obtain information about their future plans. When the captured rebels refused to give away their secrets, the unit commander ordered his soldiers to brutally torture them until they confessed. At first they were reluctant, but after being ordered several times, the military soldiers brutally tortured the rebels until they confessed.

Importantly, when considering the scenario above, one could focus either on the commander or on the military troops as the primary moral agent responsible for torturing the rebel soldiers. Taking advantage of this feature, we asked participants whether they agreed or disagreed either with an active sentence, focused on the forcer, or a passive form of the same sentence, focused on the forcee:

Forcee-focused: *The **military soldiers** were forced to brutally torture the rebels by the unit commander.*

Forcer-focused: *The **unit commander** forced his soldiers to brutally torture the rebels.*

¹ For instance, consider the report of the torture of Isa Abdullah Isa in Bahrain. ‘On the day of his arrest, Isa Isa was put in the corner of a room at the CID compound where guards formed a semi-circle around him, punching and kicking him. An officer of superior rank came into the room and told the guards to stop. He said to Isa, “Now, no one can help you. I will ask you one question and if you don’t answer it this is your last day on earth. Where is the weapon?” Isa responded, “What weapon?” The officer then ordered the guards to continue hitting Isa. Isa said that after what he estimated to be 10 minutes, he felt as if he was going to die and confessed falsely to hiding a weapon in a cemetery in the village of Sanabis’ (Human Rights Watch 2011). For the full report and details of other similar cases, see: http://www.hrw.org/sites/default/files/reports/bahrain0210webwcover_0.pdf

Note that while the focus is shifted from one agent to the other, the two sentences share the same underlying logical structure.

We hypothesized that when participants were focused on the military soldiers (“the forcee”), they would judge that they were *not* forced to torture the rebels because participants would have considered counterfactual alternatives for the military soldiers’ actions (e.g., they could have resisted the commander’s orders). Yet, we also hypothesized that, paradoxically, if participants were focused instead on the unit commander (“the forcer”), they would judge that he *did* force the military soldiers to torture the rebels because they would have considered counterfactual alternatives to the commander’s actions (e.g., he could have *not* ordered his soldiers to torture the rebels). Thus, we hypothesized that participants’ focus on either the commander or the soldiers as the primary moral agent would affect their consideration of counterfactual alternatives, resulting in a paradoxical pattern of force judgments.

Methods

Eighty participants (mean: 34 years old; 52 females) were recruited using Amazon Mechanical Turk. Participants were randomly assigned to either the forcee-focused or the forcer-focused condition. After reading the same vignette about the military commander and his soldiers, participants indicated their agreement with either the *forcee-focused* or *forcer-focused* statements. Subsequently, participants indicated their agreement with a statement to which indicated that the military soldiers had ‘no other option.’ Participants reported their agreement for both questions on a 7-point Likert Scale anchored at 1 ('Completely disagree') and 7 ('Completely agree').²

Results

As hypothesized, participants agreed that the unit commander forced the soldiers to torture the rebels ($M = 5.54$) significantly more than they agreed that the soldiers were forced by the unit commander to torture the rebels ($M = 4.53$), $t(77)=2.78$, $p=.007$). Moreover, participants in the forcee-focused versus forcer-focused conditions delivered different judgments of whether the military soldiers had no other option. Participants were less likely to agree the soldiers had no other option when they were focused on the soldiers (forcee-focused) ($M = 3.4$) than when they were focused on the commander (forcer-focused) ($M = 5.2$), $t(77)=4.49$, $p<.001$ (Fig. 1).

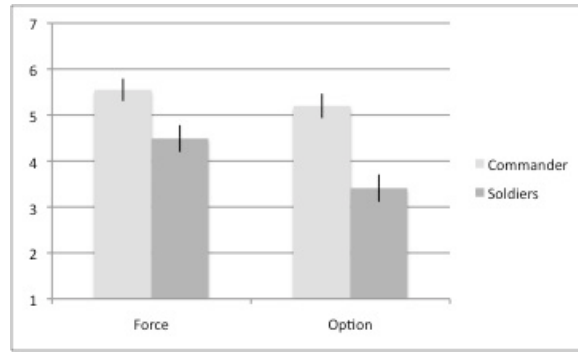


Figure 1: Participants’ agreement ratings by focus (Commander vs. Soldiers) for both statements concerning *force* (left) and *no other options* (right) on a scale from 1 ('Completely disagree') to 7 ('Completely agree').

Judgments of force and judgments of whether the military soldiers had no other options were themselves significantly correlated such that those who agreed that the military soldiers were forced also tended to agree that they had no other option ($r = .65$, $p < .001$). Additionally, we conducted a mediational analysis to test whether, as hypothesized, judgments of having no other option mediated the effect of moral focus (forcee-focused versus forcer-focused) on judgments of force. When focus and no other option were entered simultaneously, the regression coefficient measuring the relationship between condition and option went from $\beta = .30$, $p < .01$ to $\beta = .01$, $p > .9$. A Sobel test showed that this reduction was significant, $Z = 5.16$, $p < .001$ (Fig. 2).

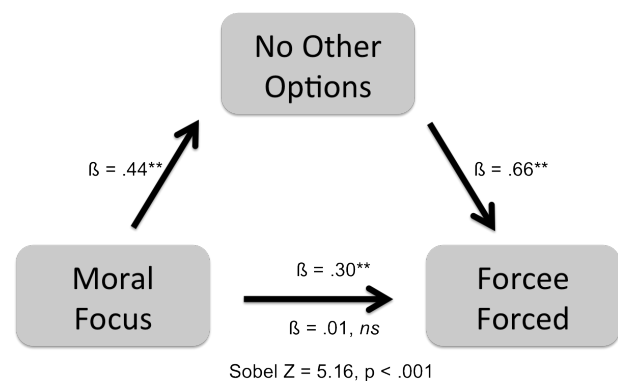


Figure 2: Standardized regression coefficients for the relationship between moral focus on forcer vs. forcee and whether the forcee was forced as mediated by absence of alternative options for forcee.

Discussion

Participants’ responses replicated the paradoxical pattern of results found in earlier studies. Here, participants were more willing to agree that the military commander forced the soldiers to torture the rebels than to agree that the soldiers were forced by the military commander to torture the rebels. Furthermore, when focused on the military commander,

² One participant was excluded from the analysis for failing a control question that determined whether participants had read the instructions.

participants were more willing to agree that the military soldiers didn't have other alternative options. In contrast, when participants focused on the military soldiers as the relevant moral agent, they were more willing to agree that the soldiers did have other alternative options. Participants' ratings of whether the soldiers had no other options mediated the effect of focus on force judgments. Thus, focus influences the extent to which alternative options available to an agent are considered, influencing judgments of force in turn.

It remains possible that participants' responses in this first study were influenced either by the rigid hierarchy in the military or by the severity of the immoral action of brutal torture. Moreover, the difference between conditions may have also been caused, not by the moral focus specifically, but rather by some other difference between the passive and active forms of the sentences. We address these alternatives in Study 2.

Study 2

Study 2 provides two contributions. First, to address the possibility that the effects observed in Study 1 are specific to severely immoral actions that take place in the context of rigid hierarchies (e.g., the military), we presented participants in Study 2 with a new vignette set at a middle school. Second, in Study 1 the passive statement was always focused on the forcee and the active statement was always focused on the forcer; Study 2 unconfounded these dimensions by introducing a new *passive* statement that focused on the *forcer*.

All participants read the following vignette:

Ralph the bully is well known at his middle school for getting what he wants by threatening the other students. Recently, Ralph almost got kicked out of school after a student named Sam told the principal that Ralph beat him up. Now Ralph has come up with a plan to get revenge by telling the principal that Sam has been sneaking into the girls' locker room. The one problem is that Ralph knows that the principal won't believe him. So Ralph tells another student named Tim that if he doesn't lie to the principal about Sam, he'll beat him up, just like he beat up Sam. Tim knows that he shouldn't lie to the principal and that Sam will probably get kicked out of school, but he tells the principal anyway because he doesn't want to get beaten up. After the principal hears that Sam has been sneaking into the girls' locker room, he expels Sam from school.

To illustrate that the difference was not due merely to the passive vs. active constructions of the sentence, we included a third sentence which was clefted such that it was focused on the forcer (Ralph the Bully) while remaining passive. Thus after reading this vignette, participants rated their agreement with one of these three logically equivalent sentences:

Passive: **Tim** was forced to lie to the principal about Sam by **Ralph the bully**.

Active: **Ralph the bully** forced Tim to lie to the principal about Sam.

Clefted: It was by **Ralph the bully** that Tim was forced to lie to the principal about Sam.

We hypothesized that agreement ratings with the passive versus active sentence constructions would replicate the previously observed paradoxical pattern: Participants would agree more that Ralph the bully forced Tim to lie to the principal but agree less that Tim was forced to lie to the principal by Ralph the bully. Further, we hypothesized that participants would agree with the clefted form of the passive sentence, showing that this effect arises specifically because of moral focus.

Methods

Seventy-five participants (mean: 38 years old; 49 females) were recruited, as in Study 1. Participants were randomly assigned to either the *passive*, *active* or *clefted* condition. After reading the vignette, participants rated their agreement on a scale from 1 ('Completely disagree') to 7 ('Completely agree').

Results

The data were analyzed using a one-way ANOVA, which revealed a main effect of condition (passive vs. active vs. clefted) on judgments of force $F(72) = 11.52, p < .001$. Post-hoc comparisons were then computed to assess differences between the three conditions, with a Bonferroni correction applied ($\alpha = .0167$). At this level, there were significant differences between both between *passive* ($M = 4.08$) and *active* ($M = 5.62$), $t(46) = 3.02, p < .01$ and between *passive* and *clefted* ($M = 6.22$), $t(49) = 4.64, p < .001$, but no significant difference between *active* and *clefted* $p > .15$ (Fig. 3).

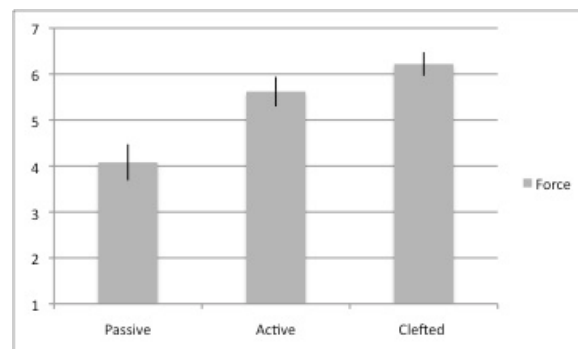


Figure 3: Participants' agreement ratings by condition for statements concerning force on a scale from 1 ('Completely disagree') to 7 ('Completely agree').

Discussion

The predicted paradoxical pattern of results again emerged. Participants agreed significantly more that ‘Tim was forced to lie to the principal about Sam by Ralph the bully,’ than that ‘Ralph the bully forced Tim to lie to the principal about Sam.’ Because this vignette did not depict an extraordinary social interaction, Study 2 provides evidence that the paradox of moral focus may actually occur in ordinary, day-to-day situations.

Moreover, despite not agreeing strongly with the statement that ‘Tim was forced to lie to the principal by Ralph the bully’, participants *did* agree that ‘It was by Ralph the bully that Tim was forced to lie to the principal about Sam.’ Finally, for the sentence constructions in which the moral focus was on the bully (*clefted* and *active*), no significant difference in agreement ratings was observed. Taken together, these results suggest it is unlikely that the observed effects are due merely to the difference between passive versus active sentence constructions. Instead, the results indicate that the mechanism underlying the paradoxical pattern of results is moral focus.

General Discussion

Situations in which more than one agent is responsible for a bad outcome allow for blame shifting – we can focus on one blameworthy agent or another. This shift in moral focus can result in a corresponding shift in the actions for which we consider counterfactual alternatives (e.g., could the forcer or the forcee have done something else to avoid this bad outcome?). Counterfactual thinking might therefore play an important role when people consider issues like force and freedom in everyday moral situations. The resulting judgments of situational constraint or force may therefore be consistent with people’s moral judgments but also reflect internal inconsistencies, as observed in the paradoxical response patterns of Studies 1 and 2.

One alternative explanation for the pattern of results is that participants may be interpreting ‘force’ not in the sense “X successfully forced Y” but as “X was *attempting* to force Y.” If this is right, then it may be the case that “X forces Y to do p” *does not* simply entail that “Y is forced to do p”. While ‘force’ may occasionally be interpreted in this way (i.e. given a performative reading), it is unlikely that the complete pattern of results can be explained on this interpretation. First, note that in each scenario, X did *successfully* force Y to do p. Second, previous research has revealed that this paradoxical pattern disappears when the action in question is morally neutral. Consider again the case of the ship captain and the sailor. If the ship captain orders the sailor to throw cargo (and not passengers) overboard, participants simply agree both that the captain forced the sailor to throw the cargo overboard and that the sailor was forced to throw the cargo overboard. Thus, alternative explanations that rely on the performative versus success distinction are unlikely in that they must assume that participants systematically relied on a performative reading for statements of the form: ‘X forced Y

to do p’ and systematically relied on a success reading for statements of the form ‘Y was forced to do p’ but only for immoral actions and despite the fact that the forcing was always successful. Thus, the overall body of evidence suggests that participants deliver inconsistent responses due specifically to moral focus and counterfactual reasoning.

Such ‘errors’ in moral reasoning fit with the literature on focusing bias more generally, which has shown that focusing effects can lead to logical errors in reasoning and decision-making across a number of contexts (Legrenzi, et al., 1993). Of course, unlike logical reasoning, moral reasoning may not generate ‘correct’ or ‘incorrect’ answers. Nevertheless, participants’ paradoxical response pattern is clearly *internally* inconsistent. This point should be especially clear when one considers these results as only two of a growing number of similarly structured studies, which have elicited the same paradoxical pattern. For example, participants in a previous study read about a chief of surgery at a hospital who orders a subordinate doctor to prescribe a drug that kills a patient. Participants judged that the chief did force the doctor to prescribe a medicine (mean judgment: 5.2 on a 1 to 7 Likert scale) but judged that the doctor was *not forced* to do prescribe the medicine (mean judgment: 3.8) (Young & Phillips 2011). Moreover, pilot data suggest that the paradoxical pattern persists even in a within-subject design.³

To what extent do such paradoxical responses occur? One possibility is that such effects occur only within the context of strict power hierarchies, e.g., military commanders and subordinate soldiers, bullies and their peers. Another possibility is that such effects are possible within a wide range of social situations. As we were specifically interested in judgments of force, we relied on situations in which there is a background power hierarchy, which is especially conducive for one person to force another. Future work should explore a broader range of relationships and situations.

Importantly, the present studies may be understood as part of a broader research program, which has suggested that moral judgments influence a whole host of other seemingly non-moral judgments. Moral judgments have been shown to influence judgments of causation (Hitchcock & Knobe, 2009; Knobe & Fraser, 2008), force (Phillips & Knobe, 2009), happiness (Phillips, Misenheimer & Knobe, 2011), intent (Knobe, 2003), knowledge (Beebe & Buckwalter, in press), and counterfactual thinking (Branscombe et al., 1996). We suspect that similarly inconsistent results may occur for these other types of non-moral judgments as well, even outside of situations in which there is a power hierarchy. For instance, consider a situation in which two

³ Participants judged that the captain forced the sailor to throw the passengers overboard (mean: 4.84) more than they judged that the sailor was forced to throw the passengers overboard (mean: 4.16; $t(82)=-3.14$ $p=0.002$). This difference, however, was driven by a subset of the 83 participants: 29 subjects showed the predicted difference.

agents jointly cause a bad outcome. Whichever agent receives more moral focus may be judged as more causally responsible for the bad outcome. Furthermore, shifting the moral focus from one agent to another may result in internally inconsistent patterns of responses, e.g., X rather than Y caused p versus Y rather than X caused p. These predictions should be investigated in future research.

Recent research has also sought to target the nature and boundaries of the impact of moral judgment on non-moral cognition. Do effects such as this reflect a genuine role for morality in concepts such as force (Knobe, 2010; Pettit & Knobe, 2009; Phelan & Sarkissian, 2008; Wright & Bengson, 2009)? Or do they provide evidence for another instance of motivated moral reasoning (Adams & Steadman, 2004; Alicke, 2008; Nadelhoffer, in press)? While this may be, at its core, a philosophical rather than empirical question, the current study seeks to move this debate forward by (1) revealing that moral evaluations can lead to paradoxical patterns of judgments even in ordinary moral situations and (2) suggesting a specific underlying mechanism for the influence of moral judgment on non-moral cognition. Focusing on immoral actions may lead people to imagine counterfactual alternatives to those specific actions and, as a result, change downstream judgments such as whether or not an agent was ‘forced.’ We leave the question of whether this same mechanism can also explain the influence of morality on other non-moral domains for future research.

Acknowledgements

We would like to thank Joshua Knobe and Zoltán Szabó for their helpful discussion on a preliminary version this research.

References

- Adams, F., & Steadman, A. (2004). Intentional action and moral considerations: Still pragmatic. *Analysis*, 64, 268-276.
- Alicke, M. (2008). Blaming badly. *Journal of Cognition and Culture*, 8, 179-186.
- Aquinas, T. (1273/1920). *The Summa Theologica of St. Thomas Aquinas*, Second and revised edition. (Fathers of the English Dominican Province, Trans.).
- Aristotle. (2002). *Nicomachean ethics*. (S. Broadie & C. Rowe, Trans.). Oxford: Oxford University Press.
- Beebe, J., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, 25, 474-498.
- Branscombe, N., Owen, S., Garstka, T., & Coleman, J. (1996). Rape and accident counterfactuals: Who might have done otherwise and would it have changed the outcome? *Journal of Applied Social Psychology*, 26, 1042-1067.
- Darley, J. M., & Shultz, T. R. (1990). Moral rules - Their content and acquisition. *Annual Review of Psychology*, 41, 525-556.
- Guglielmo, S., Monroe, A., & Malle, B. (in press). At the heart of morality lies folk psychology. *Inquiry*.
- Hart, H. L. A. (1968). *Punishment and responsibility*. Oxford: Oxford University Press.
- Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, 11, 587-612.
- Human Rights Watch. *Torture redux: The revival of physical coercion during interrogations in Bahrain*. New York. 2001
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.
- Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315-329.
- Knobe, J., & Fraser, B. (2008). Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (pp. 441-444). Cambridge, MA: MIT Press.
- Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition*, 49, 37-66.
- Machery, E. (2008). The folk concept of intentional action: philosophical and experimental issues. *Mind & Language*, 23, 165-189.
- Mallon, R. (2008). Knobe versus Machery: Testing the trade-off hypothesis. *Mind & Language*, 23, 247-255.
- N’gbala, A., & Branscombe, N. (1997). When does action elicit more regret than inaction and is counterfactual mutation the mediator of this effect? *Journal of Experimental Social Psychology*, 33, 324-343.
- Nadelhoffer, T. (in press). Intentional action and intending: Recent empirical studies. *Philosophical Psychology*.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs* 41, 663-685.
- Pettit, D., & Knobe, J. (2009). The pervasive impact of moral judgment. *Mind & Language*, 24, 586-604.
- Phelan, M., & Sarkissian, H. (2008). The folk strike back: Or, why you didn’t do it intentionally, though it was bad and you knew it. *Philosophical Studies*, 138, 291-298.
- Phillips, J., & Knobe, J. (2009). Moral judgments and intuitions about freedom. *Psychological Inquiry*, 20, 30-36.
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100, 283-301.
- Wright, J., & Bengson, J. (2009). Asymmetries in Judgments of Responsibility and Intentional Action. *Mind & Language*, 24, 24-50.
- Young, L., & Phillips, J. (2011). The paradox of moral focus. *Cognition*, 119: 166-178.